Filippo Catani

Landslide detection by deep learning of non-nadiral and crowdsourced optical images

Abstract The recent development of mobile surveying platforms and crowdsourced geoinformation has produced a huge amount of non-validated data that are now available for research and application. In the field of risk analysis, with particular reference to landslide hazard, images generated by autonomous platforms (such as UAVs, ground-based acquisition systems, satellite sensors) and pictures obtained from web data mining are easily gathered and contribute to the fast surge in the amount of non-organized information that may engulf data storage facilities. Therefore, the high potential impact of such methods is severely reduced by the need of a massive amount of human intelligence tasks (HITs), which is necessary to filter and classify the data, whatever the final purpose. In this work, we present a new set of convolutional neural networks (CNNs) specifically designed for the automated recognition of landslides and mass movements in non-standard pictures that can be used in automated image classification, in supporting UAV autonomous guidance and in the filtering of data-mined information. Computer vision can be of great help in fostering the autonomous capability of intelligent systems to complement, or completely substitute, HITs. Image and object recognition are at the forefront of this research field. The deep learning procedure has been accomplished by applying transfer learning to some of the top-performer CNNs available in the literature. Results show that the deep learning machines, calibrated on a relevant dataset of validated images of landforms, may supply reliable predictions with computational time and resource requirements compatible with most of the UAV platforms and web data mining applications in landslide hazard studies. Average accuracy achieved by the proposed methods ranges between 87 and 90% and is consistently higher than that obtained by general-purpose state-of-the-art image recognition convolutional neural networks. The method can be applied to early warning, vulnerability assessment, residual risk estimation, model parameterisation and landslide mapping. Specific advantages will be the reduction of the present limitations in the intelligent guidance of landslide mapping drones, the classification of fake news, the validation of post-disaster information and the correct interpretation of an impending change in the environment.

Introduction

The science of natural hazards, including landslides, has lately been positively impacted by the quick growth of remote sensing and crowdsourced platforms such as satellites, UAVs, social networks, sensor networks and public online data storages. For example, the usage of air- and UAV-borne sensors has gained a notable relevance due to the concurrent effect of price-lowering and quality gain in rotors, structure materials, power systems, onboard computing power and sensors (Giordan et al. 2018; Rossi et al. 2018) and to the multiplication of UAV-based applications in most of the research and application fields of science, industry and defence (Lee et al. 2017). An even stronger increase has been observed in the availability of crowdsourced information generated by data mining web resources of various type, with a special relevance of geo-tagged unclassified and potentially useful images.

In turn, this has generated an exponential surge in the amount of available data that contain large quantities of noisy and non-validated information. To be usable, such big data require automation and the support of machine learning methods for selection, classification and storage (Catani et al. 2013; Smith et al. 2017; Intrieri et al. 2017; Du et al. 2019).

For the specific case of image-related data, where the information content is carried by a multi-layered digital matrix of quantitative measures in n dimensions, computer vision methods may be of great help, because they are capable of mimicking simple and repetitive human decision tasks, if suitably trained.

With reference to the specific field of landslide hazard and risk assessment, the usage of unmanned platforms has recently become almost mandatory, due to the operational flexibility, high spatial resolution, low cost, quick capability of deployment and availability of a number of new sensors that were previously unavailable as payload on such small aircrafts (Niethammer et al. 2012; Lucieer et al. 2014; Turner et al. 2015; Giordan et al. 2015; Giordan et al. 2018; Allasia et al. 2019).

On a quite different path, direct surveying is being increasingly complemented by the usage of data mining of web-related and crowdsourced information (Battistini et al. 2013; Battistini et al. 2017; Smith et al. 2017). This indirect approach provides an alternative way to explore the occurrence of hazards over large areas and backwards in time. It also allows for the collection of soft data such as damage estimation, impacts on population, reaction time during emergency and system resilience, which are fundamental for the calibration and validation of risk assessment models (Corominas et al. 2014; Uzielli et al. 2015a; Uzielli et al. 2015b). Several text and semantic analysis methods exist that can be fruitfully used for the selection and classification of online news and automated positioning of events (Battistini et al. 2013; Smith et al. 2017) even though not much exists concerning the analysis of more complex data on landslides, such as photographs, multi-spectral images and multi-source web graphics.

Therefore, the exploitation of digital images is at the forefront of the research challenges and being improved at a fast rate.

Most of the advantages of both crowdsourced and UAV-surveyed imagery derive from the easiness of use and the short time required to gather historical, monitoring or mapping data (Bishop et al. 2012; Corominas et al. 2014; Chae et al. 2017; Giordan et al. 2018). However, when the sorting and classification of thousands, if not hundred thousands, of completely different images entail a repeated human intelligence task (HIT) or when an efficient drone-based survey requires a direct or indirect human control by an expert pilot, most of the advantages may be lost and strong limitations may be introduced due to many factors,

including time constrains, data formats, terrain configuration and logistics, thereby reducing applicability and extent of data collection. For such reasons, recent cutting-edge research is trying to perfect the computer vision proficiency in object recognition on the one side and the autonomous flying capabilities of drones to allow the execution of larger scale, all-terrain surveys, on the other side (Niesterowicz and Stepinski 2013; Lee et al. 2017).

The autonomous recognition and guidance capabilities of machines are challenging tasks that are being tackled by the research community in several ways (Minaeian et al. 2016; Lee et al. 2017). All of them entail, as a basic requirement, the capability of computer vision by the machine platform, for decision-making, obstacle avoidance, path adjustment and object detection.

Object detection, in particular, is a very important add-on to any autonomous system as a specific skill that supports intelligent decisions by helping the CPU in the interpretation of complex data extracted from the surrounding environment. Examples of such skills are the proficiency in object recognition from simple photographs, the autonomous extraction of flying information and the generation of additional smart data for optimizing survey operations or validating models.

In the field of landslide hazard, one of the main tasks which is devoted to drone systems is the quick survey of areas that are too large to be inspected with ground visits, yet require a detail-scale analysis. On the other hand, data mining systems can be used to collect large-scale information in real-time and back-analysis concerning cases of damage and risk assessment (Battistini et al. 2017). In both cases, the computer vision system should mostly concentrate on the capability of correctly classifying the terrain in terms of landforms, processes and effects due to the action of mass movements, while being at the same time capable of detecting the presence of elements at risk, such as buildings, structures and infrastructures. A specific challenge is linked to the fact that most of web-sourced and UAV-generated imagery for landslide studies is non-nadiral and non-standard (Minaeian et al. 2016).

There are many studies reporting on effective and accurate methods to map landslides from optical and non-optical imagery. An important review work by Evans (2012) proposes a conceptual framework for the interpretation of landforms, which is a starting point for every automated analysis to tackle multi-scale issues. Further developing the idea of landform delineation, Jasiewicz and Stepinski (2013) propose the operational concept of the geomorphons, as the basic landscape unit to be classified with the help of pattern recognition methods. Later on, the accuracy requirements on landform measurement needed by specific geomorphic analysis have been classified and discussed by several authors (Tarolli 2014; Eltner et al. 2016).

On such a basis, a relevant literature exists covering landform recognition. Examples include methods based on classical pixel-based satellite image classification (Liu and Wu 2016), super-pixel segmentation (Li et al. 2018), object-based image analysis (Drăguț and Blaschke 2006; Lu et al. 2011; Stumpf and Kerle 2011; Drăguț and Eisank 2012; Hölbling et al. 2016), combination of multi-spectral measurements with DEM-derived landform attributes such as elevation, slope, topographic position, and contributing area with watershed delineation (Mondini et al. 2011; Forzieri et al. 2012; Forzieri et al. 2013; Ciampalini et al. 2016; Du et al. 2019). An overview on such studies is provided by Scaioni et al. (2014) and, more recently, by Giordan et al. (2018). Most of the studies agree on the fact that deep learning convolutional neural networks

(CNNs) may be an optimal solution for highly flexible and powerful image classification and object recognition (Shin et al. 2016; Du et al. 2019). In general, artificial neural networks have long been successfully used to recognize specific landscape characters leading to slope instability (Lee et al. 2004; Catani et al. 2005; Ermini et al. 2005; Pradhan and Lee 2010; Yilmaz 2010; Liu and Wu 2016; Zhou et al. 2018a) or to detect anomalous displacements of rock and soil masses (Zhou et al. 2018b).

Almost all the published research concentrates on the post-processing analysis of multi-source data to apply pattern recognition and object-oriented methods for landform classification, with some of them specifically targeting mass movements. Only a few published works (Huang et al. 2011; Niesterowicz and Stepinski 2013; Lee et al. 2017), to the best of our knowledge, focus on the attempt of achieving real-time target detection for landforms or landscape scenes with computer vision. And no work at all proposes an operational method to give on-board detection capability to any intelligent system as related to mass movements.

In this paper, we propose a simple, computationally compatible deep learning classifier (LanDLC) trained for the detection and classification of specific landslide-related landforms in nadiral and non-nadiral images. The four versions of LanDLC presented in the following sections are based on the transfer learning of pre-trained general-purpose image classification convolutional neural networks that have been specifically modified towards landslide recognition.

All LanDLC versions can be fully implemented in a desktop data mining toolbox to complement existing automated context extraction and news classification applications (see, e.g. the systems described by Smith et al. (2017) and Battistini et al. (2013)). Furthermore, despite being in a prototypal stage for UAV on-board implementation, LanDLC may provide a contribution towards the objective of building self-aware drones capable of mapping landform instability and geo-hydrological hazard by independently flying over an area and targeting specific terrain features to be surveyed, positioned and stored in digital form.

Materials and Methods

Methodology

Image analysis and classification in the Earth sciences and in the broader field of remote sensing has a long and successful history that has now undergone a huge step forward due to the capability of computers to manage and process big data with artificial intelligence methods. When dealing in particular with image classification and object recognition, the highest performances, at the present state of the art, are those provided by deep learning tools, such as CNNs, that are capable of performing classification tasks directly from images rather than by using pre-selected features of them (Krizhevsky et al. 2012; He et al. 2015a; Shin et al. 2016). A CNN combines multiple nonlinear processing layers using simple elements working in parallel. The layers are interconnected by nodes and each layer uses the previous layer's output as input. Differently from other machine learning systems, CNNs may autonomously extract features from images, use them in the learning process, select only the most useful of them (activations) and then implement a highly accurate object recognition machine, based on a set of training images (Russakovsky et al. 2015; Shin et al. 2016).

However, the training of a deep CNN with tens or hundreds of layers over a large data set of images is a non-trivial task that

requires a huge computational effort preceded by a similarly large undertaking that is necessary for collecting and labelling hundred thousands, if not millions, of training images (Russakovsky et al. 2015). As an example, the general-purpose image classification CNN AlexNet (Krizhevsky et al. 2012), which is quite simple and has only eight learnable layers, uses 61 million parameters trained over several million labelled images. Luckily, such heavy duties have already been accomplished by the leading computer vision research groups for general-purpose image analysis and can be fruitfully exploited as a starting base for a much simpler process of specialized training called transfer learning (Shin et al. 2016).

Transfer learning consists in the specialized training of a subset of the deepest layers of a CNN that has been already trained for similar, but more general, classification purposes. An entire class of such public-domain CNNs exists offering various levels of flexibility, complexity and accuracy, depending on the user requirements. By picking one of such pre-trained, non-specialized networks, it is possible to substitute the deepest layers and retrain them to fit very specialized tasks such as the classification of landforms characterized by mass wasting and landslides. Because most of the classification capability of the network has already been obtained, transfer learning can be performed with a relatively small number of specialized images belonging to the target category. Furthermore, the usage of a general-purpose object recognition CNN strongly enhances the capability of detecting single objects set against a complex background which may include other landscape features such as trees, buildings, clouds, roads, people and animals.

In this paper, we selected four among the best performing CNN architectures for image recognition and object detection as related to the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al. 2015) and tested them by transfer learning on a dataset of labelled landscape images containing verified landforms belonging to five categories ('landslide', 'scree deposit', 'rock cliff', 'alluvial fan' and 'slope without mass movements').

The choice of the five categories is based on the following reasons: landslides are the target object for the detection system we want to develop; scree deposits, alluvial fans and rock cliffs are typical landforms that can be erroneously classified as landslides and that, therefore, have to be discriminated from them; finally, 'slope without mass movements' is the label assigned to any image in the dataset where none of the previous categories is present, according to a careful expert-based selection process. Most of the selected 'slope without mass movements' images, however, purposefully contain objects that can be mistaken for slope processes, such as mid-slope roads, buildings, cultivated fields, retaining walls and rivers. This should contribute to a more effective training of the network and decrease the degree of overfitting (Zhou et al. 2016; Lee et al. 2017).

The four pre-trained CNNs tested in this work derive from the successors of the AlexNet architecture (Krizhevsky et al. 2012) and its derivations. All of them are on the Pareto frontier and Pareto-efficient in the domain accuracy versus prediction time (Fig. 1). Any set of non-dominated solutions, being chosen as optimal, can be defined as Pareto-efficient if no objective can be improved without sacrificing at least one other objective. On the other hand, a solution ζ^* is referred to as dominated by another solution ζ if, and only if, ζ is equally good or better than ζ^* with respect to all objectives. In such terms, the chosen CNN architectures are state-of-the-art at the present stage and excelling in the combination of accuracy and computational efficiency.
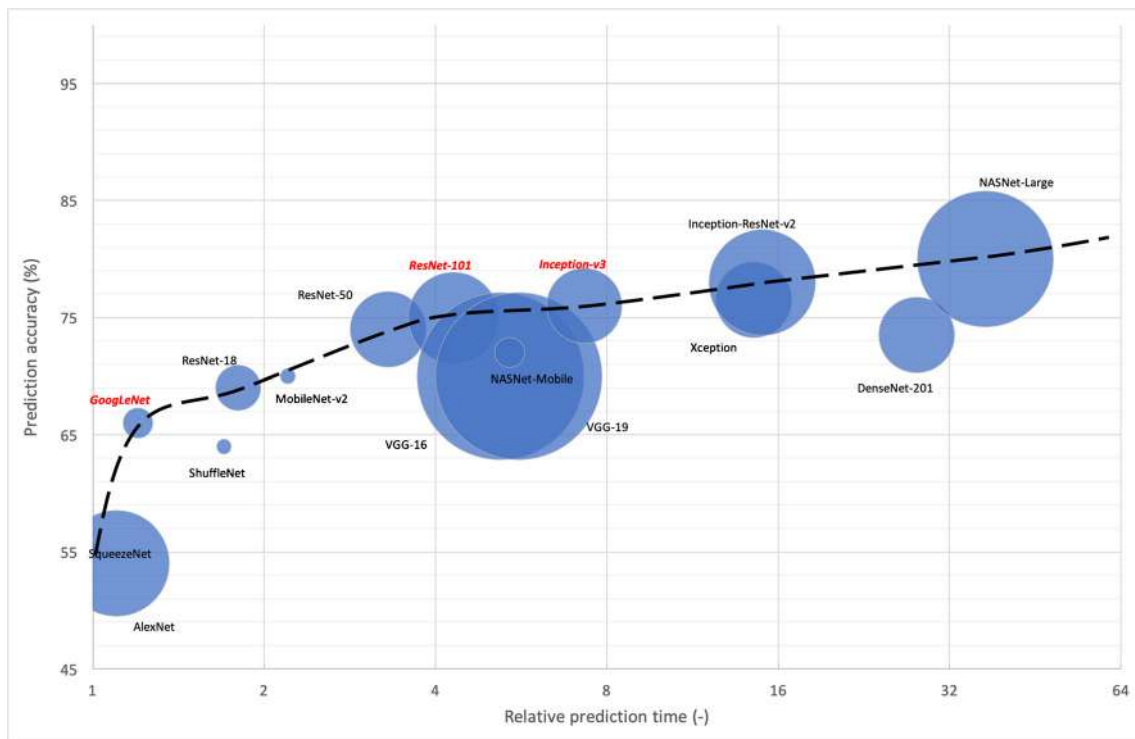

Fig. 1 Position of some popular CNNs along the Pareto frontier (dashed line) in terms of accuracy vs prediction time with respect to the ImageNet database. The four CNNs used in this paper are highlighted in bold red. Data from Mathworks (www.mathworks.com)

Table 1 CNN architecture, characteristics and parameter domains used for the training tests in the optimization procedure. For each CNN, the number of learnable layers is shown together with the total number of layers and the number of parameters. The image input size is also reported

CNN name	Original architecture	Input resolution (pixels)	Learnable layers	Total layers	Parameter no.
Go-LanDLC	GoogLeNet	224x224	22	144	7.0×10^6
GP-LanDLC	GoogLeNet.Places365	224x224	22	144	7.0×10^6
Re-LanDLC	ResNet.101	224x224	101	347	44.6×10^6
In-LanDLC	Inception.v3	299x299	48	315	23.9×10^6
Mini-batch size (−)	10, 20, 30				
Initial learning rate (−)	1.0×10^{-5}, 5.0×10^{-5}, 1.0×10^{-4}, 5.0×10^{-4}, 1.0×10^{-3}, 5.0×10^{-3}, 1.0×10^{-2}, 5.0×10^{-2}				
Momentum (−)	0.0, 0.2, 0.4, 0.6, 0.8, 1.0				

They are as follows: GoogLeNet, compact and fast with a large degree of flexibility and a good overall accuracy (Szegedy et al. 2015a); GoogLeNet-Places365 (Zhou et al. 2016), a modified version of GoogLeNet specifically oriented towards the classification of the scene rather than single objects; ResNet.101, a 101-layer CNN with improved training curve based on residual learning (He et al. 2015a); and Inception.V3, possibly the latest state-of-the-art open-source network for classification of multi-purpose images in near real time (Szegedy et al. 2015b). While the two GoogLeNet-derived CNNs are compact and fast, ResNet and Inception.v3 are more accurate at the expenses of requiring more computing power and being less compact in terms of potential UAV and robot implementation. Architectures with potentially higher accuracy than Inception.v3 require a prediction time more than double (Fig. 1) and have not been considered in this study due to their low suitability to operational near-real-time applications.

The transfer learning was performed by removing the classification and SoftMax layers at the end of the network structure and the learnable layers (convolutional or fully connected) just before them from the pre-trained CNNs, then by replacing them with new layers specifically designed for landform classification in five classes, as previously specified (all network architectures and specifications are provided upon request under a CC-BY-NC 3.0 licence in ONNX format).

All transfer learning and training was done in the Matlab environment (®Mathworks). Since performances in terms of accuracy, flexibility and overfitting avoidance are linked not only to the network architecture but also to training options, we performed a multiple-parameter optimization procedure based on a combination of three training regulation variables: (i) mini-batch size, that is, the size of the subset of images used for each iteration; (ii) initial learning

rate, that is, the scale of the search lag in the error minimization procedure; and (iii) momentum, that is, the adjustment factor for avoiding target misdetection in the search of function minima. The combination of all the considered parameter domains sums up to 144 different configurations for each of the four trained CNN architectures, for a total of 576 training runs for each trial. Table 1 lists the different values used for the parameter domains as well as the main characteristics of the network architectures.

At the end of the optimization cycle, the results were ranked by overall accuracy to select the best method and parameter set for the choice of the optimal CNN configuration which was thereafter compiled and executed against an external validation dataset made up by unlabelled images to simulate an actual operational application. No direct comparison in classification performances is possible between the modified CNNs here developed (LanDLC) and the original pre-trained networks (GoogLeNet, GoogLeNet.Places365, ResNet.101 and Inception.v3) because the latter do not include the 5 labels which are the target of the research (landslide, stable slope, rock cliff, alluvial fan and scree deposit). Therefore, we only compare average accuracy of original networks as reported in the literature, as visible in the Pareto frontier plot of Fig. 1, to the average accuracy obtained by the LanDLC networks.

The training and testing of all the configurations during optimization were performed on a multi-GPU platform (CUDA NVIDIA GeForce RTX 2070 with 36 processing cores) by using the Matlab Deep Learning toolbox (Mathworks) supported by the specific packages for the four CNNs chosen for the experiment (see Table 1). The best CNN obtained for each basic type has been validated against an independent data set and then saved for usage with external packages in ONNX format (https://onnx.ai) with the name highlighted in the first column of Table 1.

Table 2 Numerical consistency of labelled images across the three datasets used. Please note that figures do not consider data augmentation, adopted during training and testing

Label	Training	Test	Validation
Landslide	1980	495	403
Stable slope	2560	640	291
Scree	1025	256	100
Fan	1230	307	168
Cliff	1045	261	195

Data sets

The need for parameter optimization and architecture selection suggests that image datasets should be split in a training and a testing subset. Moreover, since overfitting is always a critical issue when re-training large, deep networks with a limited amount of data, a further independent dataset is required, for external validation (Russakovsky et al. 2015).

The images for training and testing are obtained through a combination of methods to ensure density (frequency of label representation) and diversity (high variability of appearances and viewpoints). Such

Table 3 Optimal configuration for each architecture. Overall accuracy for the combination showing best performances is also reported. Image classification time is relative to a single processor Intel Core i7 (2.7 GHz, 4 cores)

CNN name	Pre-trained architecture	Optimal ILR (—)	Optimal Mom (—)	Optimal MBS (—)	Overall accuracy (—)	Image classification time (s)	Size in ONNX format (MB)
Go-LanDLC	GoogLeNet	1.0×10^{-3}	0.8	10	0.88	0.025	23.9
GP-LanDLC	GoogLeNet-Places365	1.0×10^{-2}	0.4	20	0.87	0.025	23.9
Re-LanDLC	ResNet.101	5.0×10^{-3}	0.4	20	0.90	0.105	170.8
In-LanDLC	Inception.v3	5.0×10^{-3}	0.6	20	0.90	0.030	87.5

datasets must be selected and supervised very carefully by an expert geomorphologist to avoid labelling errors or multiple labelling.

Landforms chosen for training are landslides of various types, scales, states of activity and materials, which are representative of a large range of physiographical settings, versus slopes without mass movements ('stable slopes' in the remaining of the paper). Furthermore, the CNNs were trained to distinguish landslides from typical slope processes that can be mistaken for proper mass movements, such as rocky cliffs, scree deposits and alluvial fans.

Most of the images were collected by taking UAV and ground pictures of the relevant categories from the archives of the Civil Protection Centre of the University of Florence (CPC-Unifi) with manual and semi-automated selection methods. To increase the discriminant capability of the trained networks, the dataset was complemented by a second catalogue, generated by data mining image search engines on the web (Google Images, Bing Images and Flickr) on query words related to the main denominations of the chosen landforms and by using the web news catalogue generated by CPC-Unifi in-house system for automated search of landslide news over the period 2010–2017 (Battistini et al. 2013; Battistini et al. 2017). The data mining of no-landslide scenes was performed by looking up terms such as 'hillslope', 'hill' and 'landscape' and then by manually verifying them one by one, and by manually sorting through the previously mentioned image catalogues.

The combination of the two different sources of information ensures a higher diversity in the visual appearances within the dataset and allows for a more comprehensive set of non-nadiral scenes. This, in turn, should extend the capability of the trained networks towards computer vision applications and automated classification of images deriving from non-standard sources. Very often, in fact, images obtained by data mining of web resources or through automated optical camera acquisitions (mounted on drones, fixed stands or collected from non-professional photographers) are not object-centred nor clean in terms of target visibility. The inclusion of such noisy data in the training set adds more flexibility and generalization capability to the automated classification machine. As it is not possible to define a certain source for all images, with special reference to those derived from undocumented web data sources, we estimate that roughly 55% of images come from ground pictures, 35% from aerial and drone acquisitions and the remaining 10% from optical satellite images.
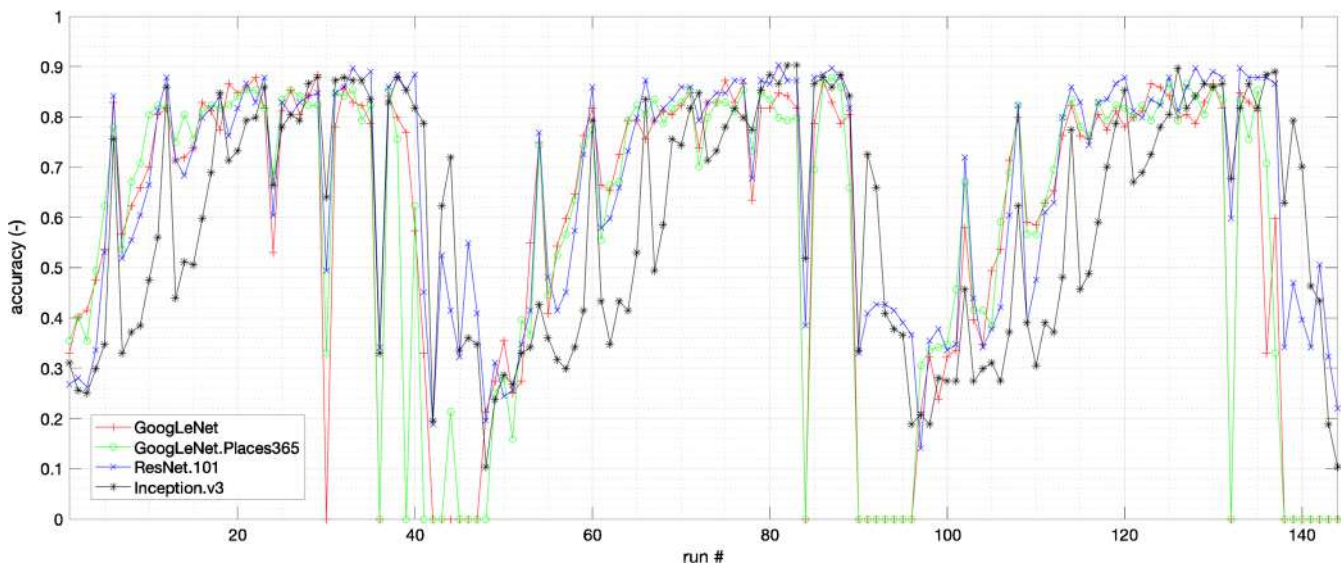

Fig. 2 Mean overall accuracy in classification for each run over the 144 combinations of optimization parameters. Points showing zero accuracy correspond to training options leading to networks with no classification capability with respect to the test set. This is often due to the adoption of wrong values of ILR
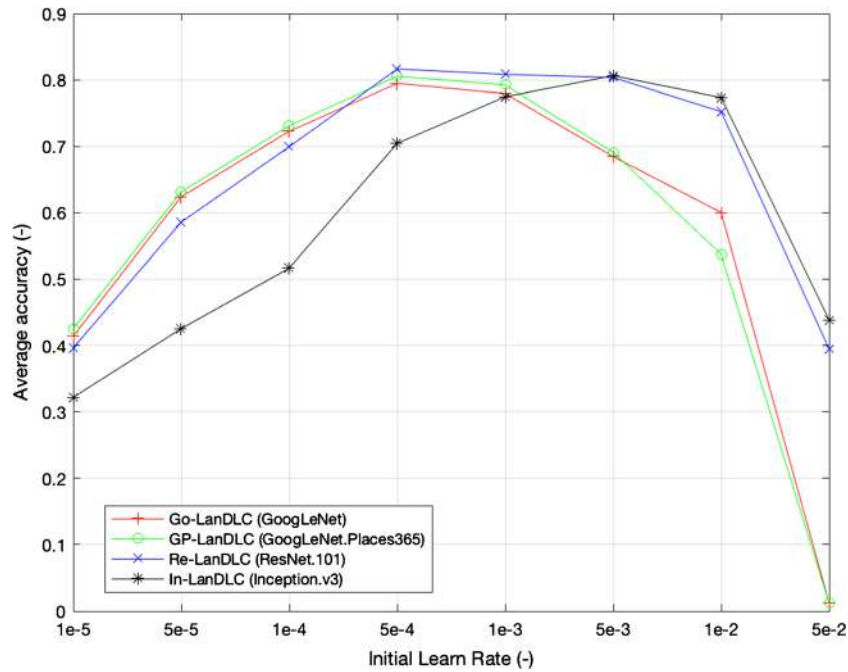
Fig. 3 Variation of the average classification accuracy of the different tested CNNs with increasing initial learning rate ILR. Typically, for large values of ILR the accuracy quickly degrades

In terms of pure numbers, after data augmentation, the dataset was split in two, with 80% (about 7900 images) of the validated images devoted to training and 20% to testing (about 2000 images). A separate validation dataset of about 1200 images was then generated by using independent non-filtered data, to simulate real-world cases of unlabelled web data mining and drone survey acquisitions. This dataset was only used after training optimization to define the actual level of reachable accuracy. In all datasets, the number of images for each class is not perfectly balanced due to the difficulty in finding suitable pictures for some specific landform types, such as scree deposits and alluvial fans. This has produced a certain degree of unbalancing in the data that has been
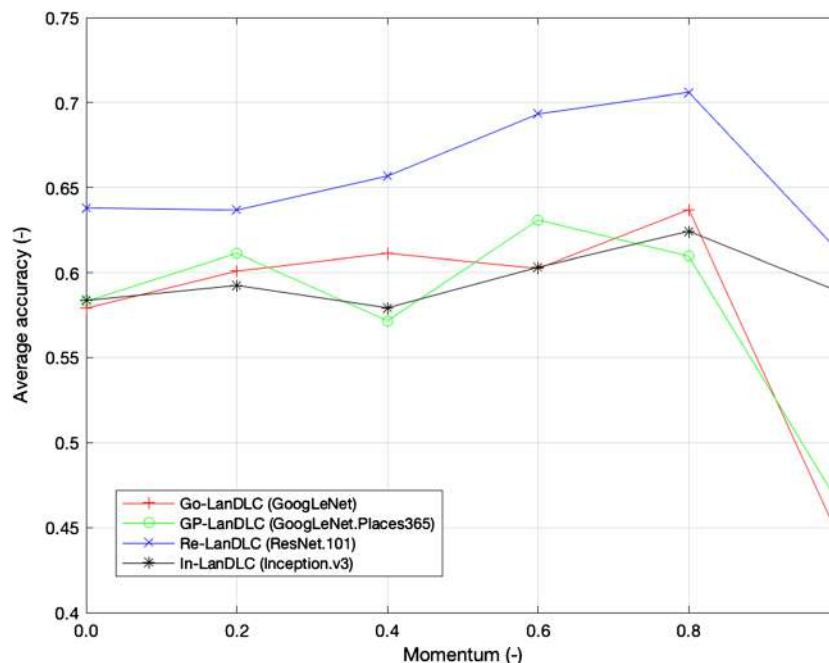

Fig. 4 Variation of the average classification accuracy of the different tested CNNs with increasing momentum

Table 4 Summary statistics on the GO-LanDLC network validation. See text for symbol explanation

Go-LanDLC optimal config Label	TP	FP	TN	FN	Precision	Recall (sensitivity)	f-score	Accuracy	Specificity	Negative predictive value	Error
Landslide	350	79	675	53	0.82	0.87	0.84	0.89	0.90	0.93	0.13
Slope	257	52	814	34	0.83	0.88	0.86	0.93	0.94	0.96	0.08
Scree	90	45	1012	10	0.67	0.90	0.77	0.95	0.96	0.99	0.05
Fan	136	11	978	32	0.93	0.81	0.86	0.96	0.99	0.97	0.04
Cliff	135	2	960	60	0.99	0.69	0.81	0.95	1.00	0.94	0.06

tackled by resorting to image augmentation techniques and by adopting suitable performance metrics (Ferri et al. 2009; Sun et al. 2009; Batuwita and Palade 2012; Branco et al. 2015). See also the Results section for details. The distribution of the used labels across the three datasets is shown in Table 2.

The density of images (number of data for each category, see Table 2) is comparable to state-of-the-art benchmarks such as the ImageNet data storage that contains over 15 million images labelled in 22,000 categories. In ImageNet, the average number of samples that are available for each category is of about 680 while in our case, each category (over the five used) has an average number of samples of about 1570, for training only. The adopted data density is even higher than that used for full training of CNNs in the ILSVRC challenge that was based on a subset of ImageNet with 1.2 million images labelled in 1000 categories, with an average density of 1200 images per category (Russakovsky et al. 2015).

During transfer learning, training and optimization, all the images were scaled to the required dimension by using an augmented image data store that combines RGB bands into a [Rx Ry 3] matrix, where Rx and Ry are, respectively, x and y image input size in pixel. During training, the augmented image data store has also been used to generate slight variations of the single images, to further increase sample density and diversity (Russakovsky et al. 2015; Zhou et al. 2016).

Results

The training of the four selected CNNs during the optimization runs shows an execution time directly proportional to the architecture complexity. For each optimization cycle, learning time was of about 2.5 min on Go-LanDLC and GP-LanDLC, 4.5 min on In-LanDLC and 6.2 min on Re-LanDLC, based on the hardware setup previously described. A series of independent post-training classification trials were also carried out on a separate hardware platform with basic computational capability, to simulate an actual operational environment on a portable platform (Intel Core i7 2.7 GHz with 4 cores). Average classification time (Table 3), using

Fig. 5 Confusion matrix for classification using Go-LanDLC (GoogLeNet) on the validation dataset. The term 'slope' is short for 'stable slope' or negative measure with respect to positive predictions

Table 5 Summary statistics on the GP-LanDLC network validation. See text for symbol explanation

GP-LanDLC optimal config Label	TP	FP	TN	FN	Precision	Recall (sensitivity)	f-score	Accuracy	Specificity	Negative predictive value	Error
Landslide	341	114	640	62	0.75	0.85	0.79	0.85	0.85	0.91	0.18
Slope	233	36	830	58	0.87	0.80	0.83	0.92	0.96	0.93	0.09
Scree	76	37	1020	24	0.67	0.76	0.71	0.95	0.96	0.98	0.06
Fan	126	61	928	42	0.67	0.75	0.71	0.91	0.94	0.96	0.10
Cliff	119	14	948	76	0.89	0.61	0.73	0.92	0.99	0.93	0.08

trained CNNs on single images was of 0.025 s for both Go-LanDLC and GP-LanDLC, 0.030 s for In-LanDLC and 0.105 s for Re-LanDLC, within Matlab. Faster CNNs are slightly less accurate, showing maximum accuracy of 0.88 (Go-LanDLC) and 0.87 (GP-LanDLC) with respect to an average 0.90 shown by the more complex Re-LanDLC and In-LanDLC. In general terms, the best performing network seems In-LanDLC, based on the Inception.v3 architecture, that has maximum accuracy equal to Re-LanDLC (based on ResNet.101) but is much faster (0.030 against 0.105 s).

For all the tested networks, accuracy seems strongly dependent on the fine-tuning of training parameters, as shown in Fig. 2 where the results of average overall accuracy are shown for each architecture and all optimization runs.

Sensitivity to the training parameters value is different for each architecture and suggests that specific optimization is needed case by case. Nonetheless, it is apparent that some parameter combinations show a general high (or low) classification accuracy for all networks.

The influence of each training parameter may be better understood by looking at the mean accuracy achieved at varying values. Figure 3 shows the variation of accuracy with increasing values of the initial learning rate (ILR), while Fig. 4 refers to the momentum (Mom).

In Fig. 3, for relatively simple architectures, based on GoogLeNet (Go-LanDLC and GP-LanDLC), the accuracy increases with ILR only up to values around 5.0×10^{-4}. After that, the accuracy diminishes with a sharp drop for ILR greater than 1.0×10^{-3}. Deeper networks, such as those based on the Inception-v3 and Resnet.101 architectures, are more robust to variations of ILR, showing a decline in prediction accuracy for values of ILR greater than 1.0×10^{-2}. In-LanDLC, in particular, exhibits an increase in the accuracy up to ILR equal to 5.0×10^{-3}.

The sensitivity to momentum appears lower (Fig. 4). Almost all architectures show a constant average accuracy until values of Mom equal to 0.8, then accuracy declines. In-LanDLC, based on
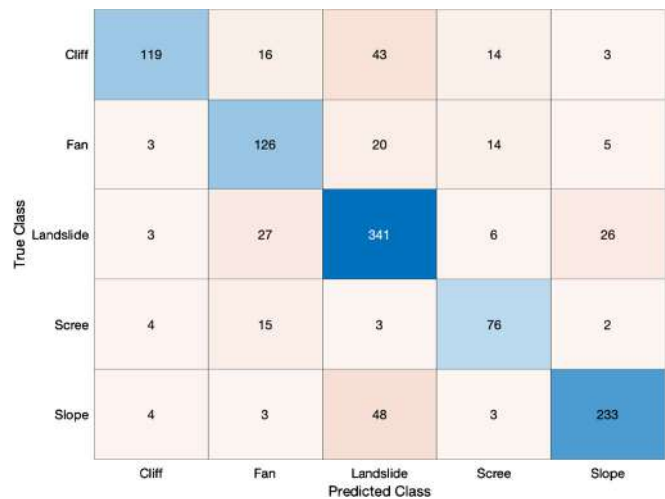

Fig. 6 Confusion matrix for classification using GP-LanDLC (GoogLeNet.Places365) on the validation dataset. The term 'slope' is short for 'stable slope' or negative measure with respect to positive predictions

Table 6 Summary statistics on the Re-LanDLC network validation. See text for symbol explanation

Re-LanDLC optimal config Label	TP	FP	TN	FN	Precision	Recall (sensitivity)	f-score	Accuracy	Specificity	Negative predictive value	Error
Landslide	356	47	724	47	0.88	0.88	0.88	0.92	0.94	0.94	0.09
Slope	258	43	823	33	0.86	0.89	0.87	0.93	0.95	0.96	0.07
Scree	89	21	1036	11	0.81	0.89	0.85	0.97	0.98	0.99	0.03
Fan	150	15	974	18	0.91	0.89	0.90	0.97	0.98	0.98	0.03
Cliff	168	10	952	27	0.94	0.86	0.90	0.97	0.99	0.97	0.03

Inception.v3, is less sensitive than other CNNs to momentum variations. The same can be said for the Re-LanDLC even though the fact that the curve is always the highest may be because, in general, the residual learning architectures are less prone to minima-seeking errors. This, however, does not mean that the Re-LanDLC is the more efficient in terms of image object recognition since computation time is much higher. Finally, the mini-batch size (MBS) is not very relevant in the tests, showing a very limited influence on overall accuracy.

The best parameter set and overall accuracy for each trained CNNs are shown in Table 3.

After training and testing, the 4 optimal configurations have been tested by running the classifier on an independent data set (see 'Materials and methods'). The results have been used to estimate the classification capability of the CNNs according to standard ranking metrics for class label data in both balanced and unbalanced samples (Ferri et al. 2009; Sun et al. 2009; Batuwita and Palade 2012; Branco et al. 2015). For each architecture and for each landform type, the following metrics have been used, where TP is the number of true positives, FP the number of false positives, TN the number of true negatives and FN the number of false negatives.

$$\text{Precision}: p = \frac{TP}{(TP + FP)}$$

$$\text{Recall (or sensitivity)}: r = \frac{TP}{(TP + FN)}$$

$$\text{F-score}: f = 2 \cdot \frac{(p \cdot r)}{(p + r)}$$

$$\text{Accuracy}: \alpha = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

$$\text{Specificity}: s = \frac{TN}{(TN + FP)}$$

$$\text{Negative Predictive Value}: npv = \frac{TN}{(FN + TN)}$$

$$\text{Error}: \epsilon = \frac{(FP + FN)}{(TP + TN)}$$

Precision (p) is a measure of the robustness towards false positives. Recall (r or sensitivity) summarizes how well positive cases are predicted accounting for the robustness towards false negatives. The f-score combines p and r into a single score. Accuracy (α) is an overall measure of correct answers with respect to total answers. Specificity (s) refers to the capability of predicting negative values against false positives. The negative predictive value (npv) measures the relative importance of false negatives. Error (ε) is the complement of accuracy and should be minimized.

The results, highlighted in the following, clarify that, in terms of predictive performance, the 4 CNN architectures behave quite differently, for each separate type of landform. The simplest and fastest CNNs, based on GoogLeNet, offer acceptable classification capability when left with the original pre-training (Go-LanDLC) and poor performances when pre-trained with the scene's dataset of Places.365 (GP-LanDLC). In Table 4, the summary statistics for the Go-LanDLC network shows that the recognition of landslides (i.e. the main target of the study) is acceptable compared to Pareto frontier averages (Fig. 1) with precision of 0.82, accuracy of 0.89 and error of 0.13. Rather good is the capability to classify alluvial fans ($p = 0.93$, $\alpha = 0.97$) and stable slopes ($p = 0.83$, $\alpha = 0.93$) as well. There is, however, a poor capability in the correct

Fig. 7 Confusion matrix for classification using Re-LanDLC (ResNet.101) on the validation dataset. The term 'slope' is short for 'stable slope' or negative measure with respect to positive predictions

classification of scree deposits (precision $p = 0.67$ and F1 score $f_1 = 0.77$). Rock cliffs show a very low number of false positives but an unsustainable number of false negatives ($p = 0.99$, $r = 0.69$). The complete results of validation for Go-LanDLC are presented in the confusion matrix of Fig. 5.

When trained on scene pictures from the Places.365 database, GoogLeNet does not improve. In fact, in Table 5, the statistics of validation for GP-LanDLC shows a quite poor classification power towards landslides ($p = 0.75$, f-score $= 0.79$, $\varepsilon = 0.18$). Even poorer is the performance with respect to scree deposits ($p = 0.67$, $r = 0.76$, f-score $= 0.71$), alluvial fans ($p = 0.67$, $r = 0.75$, f-score $= 0.71$) and rock cliffs ($p = 0.89$, $r = 0.61$, f-score $= 0.73$).

This behaviour may appear as unexpected, due to the fact that GoogLeNet.Places365 has been pre-trained on a dataset of images representing places so as to be able to classify specific site typologies. A more careful analysis, however, reveals that this pre-training is very good when the target is a set of classes representing generic places or place names, but it may be quite inefficient if the objective is to recognize an object inside a complex landscape. For example, GoogLeNet.Places365 is capable of distinguishing whether a picture is representing a classroom or a library but cannot tell whether the objects 'book' or 'computer' are present in the picture itself. This, conversely, is typically feasible by resorting to GoogLeNet with standard training. In our specific case, we are training a CNN that must detect the presence of complex objects merged in a background that is not relevant. In other words, we want to be able to recognize a landslide (or another similar landform) that is overlapping (or overlapped by), e.g. a road, a series of buildings, some vineyard lines, a parking lot or a standing passer-by. This specific task is evidently better accomplished by Go-LanDLC rather than GP-LanDLC. The complete results of the validation for GP-LanDLC are presented in the confusion matrix

of Fig. 6. In any case, GoogLeNet-based CNNs are compact and fast (ONNX size of about 24 MB and average image classification time of 0.025 s).

The increase of architecture complexity in terms of number of learnable layers appears to boost overall performances. The most advanced CNN used, Re-LanDLC based on ResNet.101, a convolutional neural network with 101 layers and residual learning, improves landslide detection with $p = 0.88$, $r = 0.88$ and $\alpha = 0.92$ (Table 6). Moreover, it strongly enhances the capability to classify scree deposits ($p = 0.81$, $r = 0.89$ and $\alpha = 0.97$), alluvial fans ($p = 0.91$, $r = 0.89$ and $\alpha = 0.97$) and rock cliffs ($p = 0.94$, $r = 0.86$ and $\alpha = 0.97$). This is done at the expenses of compactness (170.9 MB) and prediction time (0.105 s). The complete results of validation for Re-LanDLC are presented in the confusion matrix of Fig. 7.

The CNN based on Inception.v3 (In_LanDLC) seems to represent a good compromise in terms of cost-benefit ratio given the fact that it is more compact (87.5 MB) and faster (average image prediction time of 0.030 s) than Re-LanDLC. As highlighted in Table 7, overall classification proficiency is still high, with landslide classification figures that are actually higher than with Re-LanDLC ($p = 0.93$, $r = 0.87$ and $\alpha = 0.93$). The same applies for alluvial fans ($p = 0.90$, $r = 0.92$ and $\alpha = 0.97$) and rock cliffs ($p = 0.91$, $r = 0.93$ and $\alpha = 0.97$). The only exception is the detection of scree deposits, with slightly lower figures, mainly concerning the number of false positives ($p = 0.74$, $r = 0.97$ and $\alpha = 0.97$). The complete results of validation for In-LanDLC are presented in the confusion matrix of Fig. 8.

Some examples of image classification are shown in the following, with the purpose of visually describing results and typical errors as compared to actual landscape components. In all figures, the classification is reported along with the membership likelihood in percentage. Classes are indicated by short terms where the term

Table 7 Summary statistics on the In-LanDLC network validation. See text for symbol explanation

In-LanDLC optimal config Label	TP	FP	TN	FN	Precision	Recall (sensitivity)	f-score	Accuracy	Specificity	Negative predictive value	error
Landslide	349	28	726	54	0.93	0.87	0.89	0.93	0.96	0.93	0.08
Slope	252	24	842	39	0.91	0.87	0.89	0.95	0.97	0.96	0.06
Scree	97	34	1023	3	0.74	0.97	0.84	0.97	0.97	1.00	0.03
Fan	155	18	971	13	0.90	0.92	0.91	0.97	0.98	0.99	0.03
Cliff	182	18	944	13	0.91	0.93	0.92	0.97	0.98	0.99	0.03

'slope' is short for 'stable slope' and has, as previously mentioned, the significance of any picture in which the CNN detector does not recognize one of the four trained landforms ('cliff', 'fan', 'landslide' and 'scree').

In Fig. 9, a selected sample of images classified by the Go-LanDLC algorithm is depicted to highlight a typical behaviour. The CNN correctly identifies all the features with some uncertainties in ascribing the coastal cliff in image (g) and the debris-flow fan in image (h). This indecision may be due to the ambiguities that the two images represent also to a skilled human expert. The coastline, in fact, may equally represent a rock cliff or a landslide scar, depending on the level of accuracy and classification choices. The debris flow is in effect dominated by the alluvial fan that it generates, and the error is understandable. On the other hand, pictures in images (e), (f) and (i) are quite challenging but are correctly classified with a low level of uncertainty.

The Fig. 10 illustrates some cases for GP-LanDLC. The tendency to overestimate the class 'landslide', quantified by the overall precision value $p = 0.75$ in Table 5, is visible in the third image of the second row (f), where a road is flanked by an average steep slope with vegetation and some rock outcrops. The low probability (60.4%) for the class, however, may in part help to understand that the attribution is uncertain. An even worst case of false positive is the image of second column, third row (h), in which the classification algorithm is almost certain (99.6%) that the ploughed fields in the background are landslide scars. Even though there is a certain probability that the slope hosts some dormant landslides, this is not actually visible from the image and the case must be considered a false positive. Better capability is shown by GP-LanDLC in recognizing the absence of trained landforms in the low-quality image (a) and the presence of a landslide in the very confusing picture in image (g), where almost the entire image is filled by a part of the landslide body, without context or contrasting background. The capability of classifying landforms that are only partially included in images is a very useful characteristic in data mining applications and in the classification of low-altitude aerial photographs.

The examples related to the ResNet.101 network (Re-LanDLC) are reported in Fig. 11. Here, the high discriminant capability of residual training networks is highlighted by the correct classification of the landslides in images (a), (h) and (i) despite the surrounding disturbance given by buildings, people and infrastructures. There is, however, a quite serious error in the central image (e) that is misclassified (even though with some uncertainty given by the likelihood of 74.8%) as a stable slope, possibly due to the presence of vegetation on the main landslide body. The possible causes of such kind of false negatives will be discussed in the next section. A typical feature of Re-LanDLC is visible when looking at all but the central image that are classified without any uncertainty by the algorithm. This is not necessarily an advantage of the method and may generate false positives especially in the crucial distinction between landslides and stable slopes (see values of $p = 0.88$ and $p = 0.86$ in Table 6).

A similar level of detection skill is given by the In-LanDLC, based on the state-of-the-art Inception.v3 convolutional neural network. In Fig. 12 a high discriminant power is shown in images (c), (f) and (h), where, again, several disturbances are present, including internal and external factors. Quite unexpected is the

Fig. 8 Confusion matrix for classification using In-LanDLC (Inception.v3) on the validation dataset. The term 'slope' is short for 'stable slope' or negative measure with respect to positive predictions

false negative in image (g) where a possibly active landslide is completely missed (likelihood for stable slope 98.9%) possibly due to the very low colour contrast between the landside body and the surrounding slopes.

Colour contrast is a typical source of errors in CNNs for image classification that exploit only RGB optical bands. A possible improvement could be obtained by adding additional bands, such as the near infrared or the short-wave infrared, if available. This, in turn, would force a complete redesign of the CNN structure and prevent from the usage of most pre-trained architectures. This and other possible error sources will be briefly discussed in the next section.

Discussion

General considerations on CNN implementation
The results reported in the previous section seem to highlight the fact that specialized convolutional neural networks derived from transfer learning behave better than their original counterparts. This is clearly visible by comparing average accuracies reported in the Pareto frontier plot of Fig. 1 with figures of overall accuracy in Table 3. This is not unexpected and is related to the very nature of a specialized convolutional neural network. The four original CNNs used for transfer learning are general-purpose classification algorithms, capable of classifying with a relatively good accuracy thousands of different object types. This holistic aptitude is obtained at the expenses of a lesser precision and accuracy in the classification of specific features that are not included in the original training database, such as geomorphic landforms. On the contrary, the four proposed post-trained CNNs are strictly trained on the five desired target classes; therefore, they have

higher precision and accuracy on that specific task but are not usable for general-purpose scene or object classification.

The best classification results are obtained by using In-LanDLC, based on the Inception.v3 architecture, one of the best open-source CNNs for image recognition in terms of accuracy-time trade-off. The only weak point of In-LanDLC is represented by the relatively high number of false positives ($p = 0.74$) in scree deposit detection, possibly due to the underestimation of the number of landslides and stable slopes. Still, In-LanDLC boasts the best performances in all the remaining classes, including landslides which are the ultimate target of the present study. When the overall classification power is compared to the parameters that constrain the operational implementation of the algorithm, In-LanDLC is absolutely superior to Re-LanDLC with figures that suggest that the latter should be discarded in case of UAV or drone applications. The ONNX-format size (Table 3) and, more importantly, the image classification time for Re-LanDLC are much higher than for In-LanDLC and not acceptable for real-time applications. In turn, simpler architectures based on the compact and fast GoogLeNet model may offer a better drone implementation suitability at the expenses of larger errors. In particular, Go-LanDLC seems the best option between the two, due to the acceptable overall accuracy that is coupled with a small size (23.9 MB in ONNX) and a good classification speed (about 0.025 s per image). The research of the optimal trade-off between In-LanDLC and Go-LanDLC depends on the scope of work, on the type of drone (or robot) platform to be used, on the type of sensors and on a complex set of operational parameters such as flight speed and altitude, land cover type, target type and lighting conditions.

Fig. 9 a–i Some examples of classification as given by the Go-LanDLC algorithm on the validation dataset. For each single image, the assigned class is indicated along with the class membership likelihood in percentage

Landslide classification errors and possible solutions

On a different matter, we have seen in the results section that classification errors, with particular reference to landslide false negatives (or missed alerts in risk assessment terms), are present in all the trained CNNs. As an example, Re-LanDLC, despite its overall accuracy, misses the detection of a quite large landside in the central picture of Fig. 11 (image e), possibly due to the vegetation regrowth that gives a colour like the surrounding slopes to the landslide body. The failure to detect landforms which may appear as quite clear to an expert human eye can be investigated by looking at the output of convolutional layers. Each convolution produces a quasi-random set of image modifications from which the training can extract the most relevant parameters for a multivariate analysis. Such convolutional products of the original image are called activations in the CNN literature and represent features that activate exchanges of information among layers, thus enhancing the classification power. In the following, we provide two examples of activations that lead to a wrong classification by one of the CNNs, as a basis for discussion.

A sample of the activations for the landslide of Fig. 11 e, as generated by the first convolutional layer of the CNN, is depicted in Fig. 13, along with the original image. It is clear that some of them are filters that highlight terrain texture while others are capable of delineating horizontal boundaries or enhancing colour contrast. Others yet are not producing any significant pattern, at least to the human eye.

Despite the fact that the landslide is quite discernible from the surroundings, Re-LanDLC is not capable of classifying it correctly (see Fig. 11e). The remaining three networks, instead, provide correct predictions (Table 8).

This behaviour may be explained by considering that Re-LanDLC is the only architecture, among the four chosen, that uses residual learning. Residual learning allows for a deeper network to be developed by reducing the impact of the vanishing gradient problem (Hochreiter 1998). This is done by adding to each convolutional layer's output the original (or upper layer's) learned output to produce a new input data for the next layer that includes also the source information (He et al. 2015a; He et al. 2015b). This technique limits the information degradation which is typical of classical CNNs by keeping the $(n-1)$th layer output as part of the input for the $(n+1)$th layer. Consequently, it is possible to build efficient very-deep CNNs with a low degree of horizontal complexity. Such deep and narrow networks are very powerful in image recognition (ResNet.101 has won several ILSVRC challenges) as it is well accounted also in the performance indicators of our
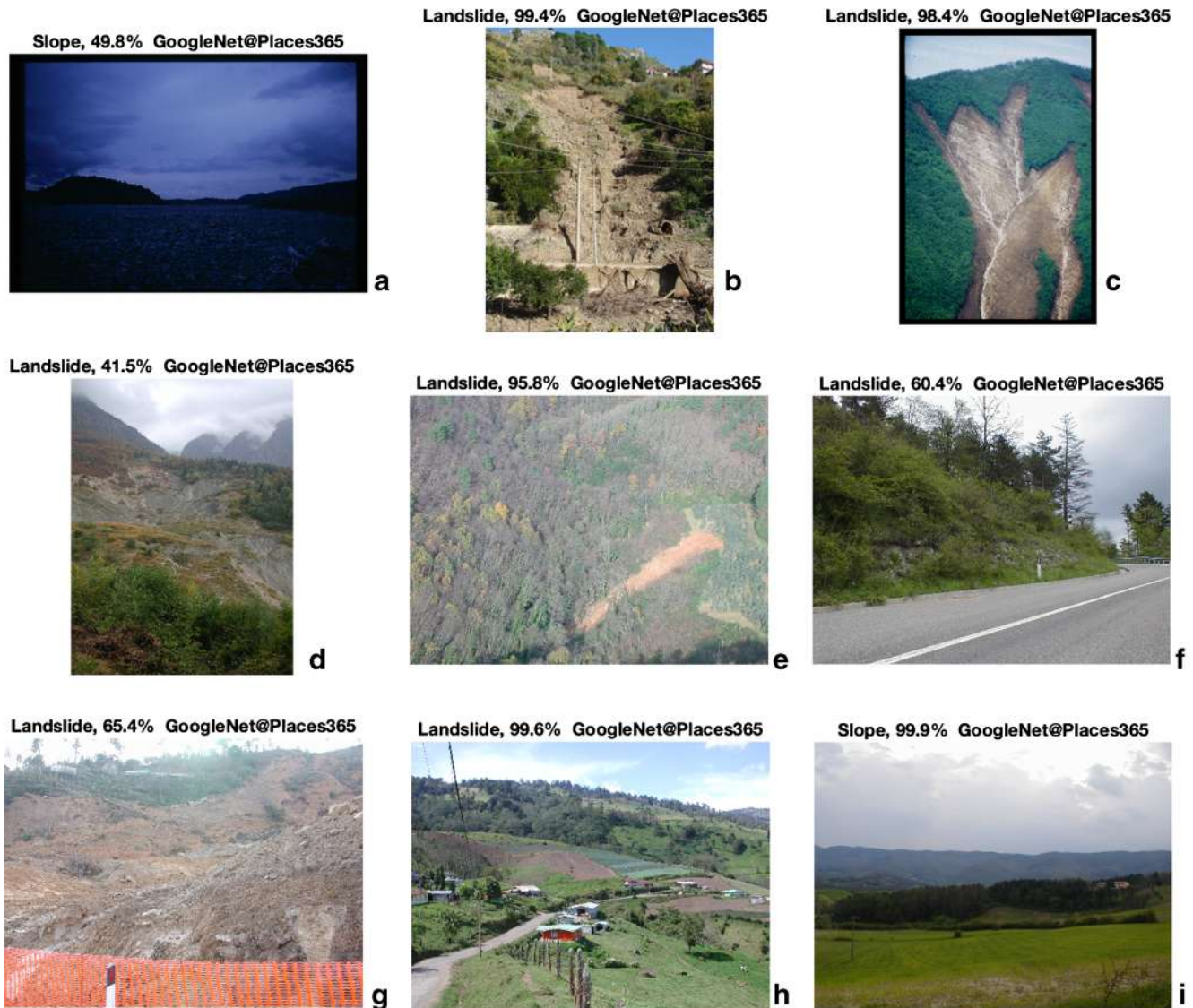
Fig. 10 a–i Some examples of classification as given by the GP-LanDLC algorithm on the validation dataset. For each single image, the assigned class is indicated along with the class membership likelihood in percentage

modified version Re-LanDLC (Table 6). However, in some specific cases, such as the one of Fig. 11 e, the residual learning may retain, in the weighting scheme, a landscape feature which is confusing rather than useful for correct classification. The same error may not occur in non-residual networks that, on the contrary, discard previous information before going deeper to the next convolutional layer.

A different case is the one depicted in Fig. 14, where a stable slope with terracing and scattered vegetation is analysed. This time, Re-LanDLC wrongly detects the presence of a landslide, thus producing a false positive (less dangerous in terms of risk assessment than a false negative). The remaining CNNs, despite some uncertainty with scree deposits and landslides, correctly identify the slope as stable (Table 9). Image activations are again the same for each network but the way each one processes the parameters emerging from them is different. In this specific case, a possible explanation of the uncertainties in the correct classification is the seeming bulging feature in the mid-right of the image generated by some of the activations. It is possible that a fine-tuning of the weighting scheme for some of the convolutional products could enhance the final prediction, thereby reducing false positives. In this case, as well, the residual learning approach of Re-LanDLC could cause a wrong weighting of the activations, by giving low scores to the most significant ones or by keeping noisy information as relevant through the residual learning technique, which inherits previous layer's convolutional outputs that may be deceiving in specific cases.

The examples reveal that a careful study of classification errors through the analysis of activations and the way the latter are propagating down within the neural network layers might reveal specific insights on image filtering methods, in order to

Fig. 11 a–i Some examples of classification as given by the Re-LanDLC algorithm on the validation dataset. For each single image, the assigned class is indicated along with the class membership likelihood in percentage

devise a set of landform-specific image convolutions and further improve the overall performances. This may be feasible only by applying a full residual learning to a partially new architecture, a task that involves a large effort in terms of image labelling and computation. This specific task is outside the scope of this paper, since it will require a specific activation analysis for each one of the images used for the training and the subsequent development of a brand-new CNN with full training. The latter, in itself, as discussed in the methodological section, would require a much larger set of labelled images for training and testing, in the order of 10^5 or larger. Provided that a similar number of landform images actually exists within publicly available resources and databases (something that, so far, remains to be verified in the first place), the correct labelling of them would only be possible by resorting to automated human intelligence tasking (HIT), such as the Amazon Mechanical Turk used for the development of GoogLeNet.Places365 (Zhou et al. 2016). However, while the human recognition of landscape scenes is a task requiring a quite common general knowledge, the analysis of landforms is a specialized task that requires a certain level of training of the HIT workers. This would surely increase the effort in terms of expected costs and time.

Automated landform detection in optical satellite images

One of the main constrains of the proposed algorithms is the small image size, which is implicit in the transfer learning technique that has been adopted to exploit pre-trained high-performance image recognition CNNs (Table 1). This limitation, however, is not relevant for the implementation of robot guidance and scene recognition and is also well compatible with frame-by-frame video analysis and crowdsourced data mining since most of the available imagery is low resolution and of limited areal coverage. Even in the case of targets with dimension much larger than the camera footprint, a simple solution is to increase flight altitude. Another, more elaborated, solution may be the automated mosaicking and resampling of drone acquisitions until a CNN-compatible scaling is obtained.
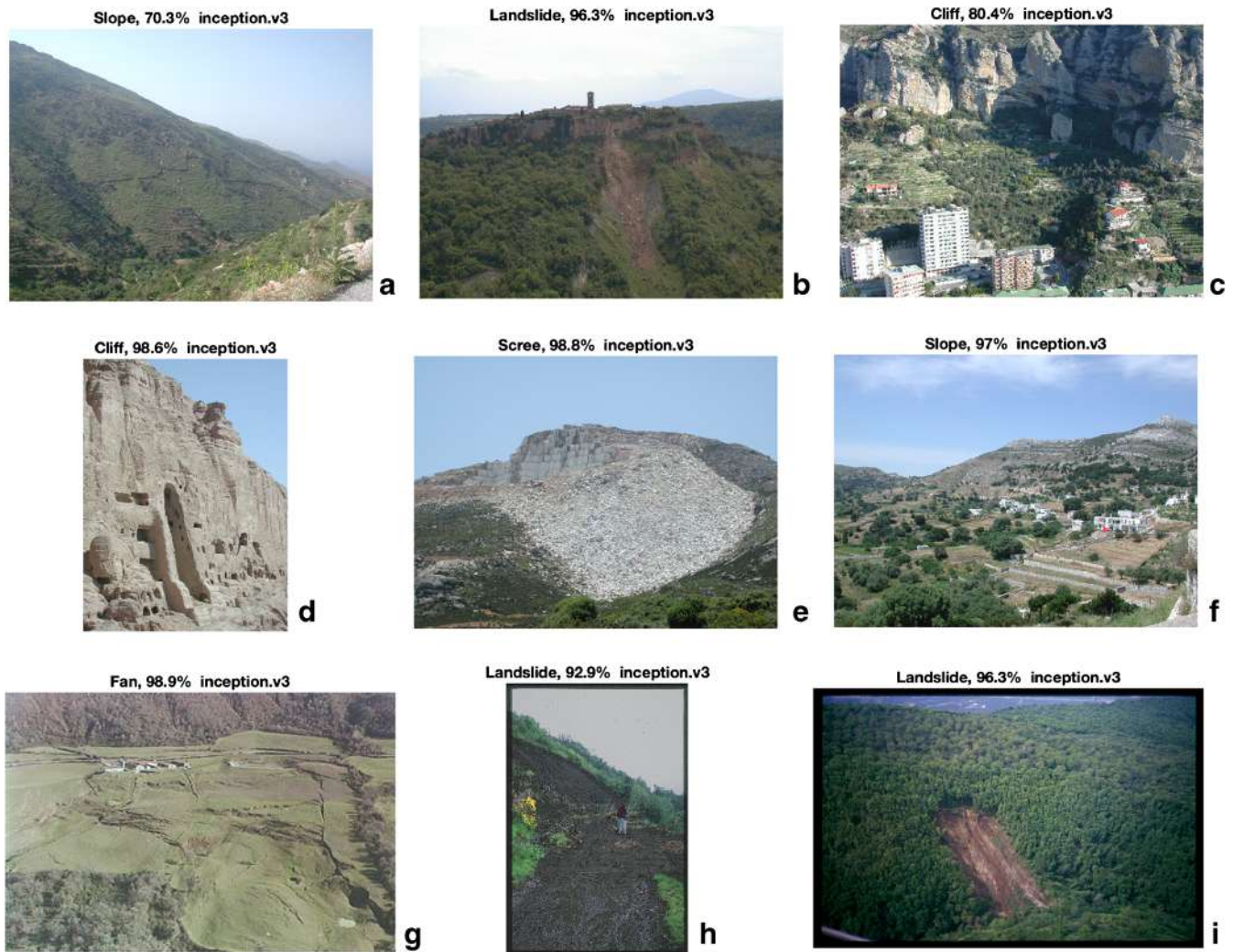
Slope, 70.3% inception.v3 — **a**
Landslide, 96.3% inception.v3 — **b**
Cliff, 80.4% inception.v3 — **c**
Cliff, 98.6% inception.v3 — **d**
Scree, 98.8% inception.v3 — **e**
Slope, 97% inception.v3 — **f**
Fan, 98.9% inception.v3 — **g**
Landslide, 92.9% inception.v3 — **h**
Landslide, 96.3% inception.v3 — **i**

Fig. 12 a–i Some examples of classification as given by the In-LanDLC algorithm on the validation dataset. For each single image, the assigned class is indicated along with the class membership likelihood in percentage

The limited image size becomes an important drawback when the image to be analysed is very large with respect to the average dimension of the target. In such a case, a downsampling of the image, to fit the required size, would completely filter out the target landforms. In the case of landslides, this may happen when trying to apply computer vision techniques to satellite optical data that cover tens of squared kilometres with a resolution of metres or tens of metres, such as Landsat 8 and Sentinel-2. On such images, having size in the order of $10^4 \times 10^4$ pixels and resolution of 10^1 m, a landslide will occupy a few pixels. Therefore, a complete image resizing to $10^2 \times 10^2$ pixels, required by the typical CNN, will definitely wipe out most of them blurring any interesting feature with the background.

A possible solution may exist, even though only in a post-processing perspective, as exemplified by some previous applications (Liu and Wu 2016). The LanDLC algorithms could be applied over a moving window, with dimensions exactly matching the required size for each adopted CNN, either in overlapping or in non-overlapping mode. For example, in Sentinel-2 optical images, multi-spectral (RGB and NIR) information is measured at a ground resolution of 10 m. That, in case of a scan size of 224 × 224 pixels (required by Go-LanDLC, GP-LanDLC and Re-LanDLC), would mean that each moving window will cover an area of 2240 × 2240 m, a dimension quite matching most of landslides, inclusive of runout. The operation should be repeated all over the satellite image for about 10^3 times in average in non-overlapping mode. That would mean, based on the classification time figures and hardware setup reported in Table 3, a total scan time in the order of 10^2 s by using the slow but powerful Re-LanDLC and of 10^1 s when using the faster and more compact Go-LanDLC. Given the fact that in post-processing operations much larger computational power can be used, such as the CUDA NVIDIA GeForce RTX 2070 with 36 processing cores used for the training or similar GPU processing units, we may expect that batch classification process chains might be implemented for large image datasets quite easily.
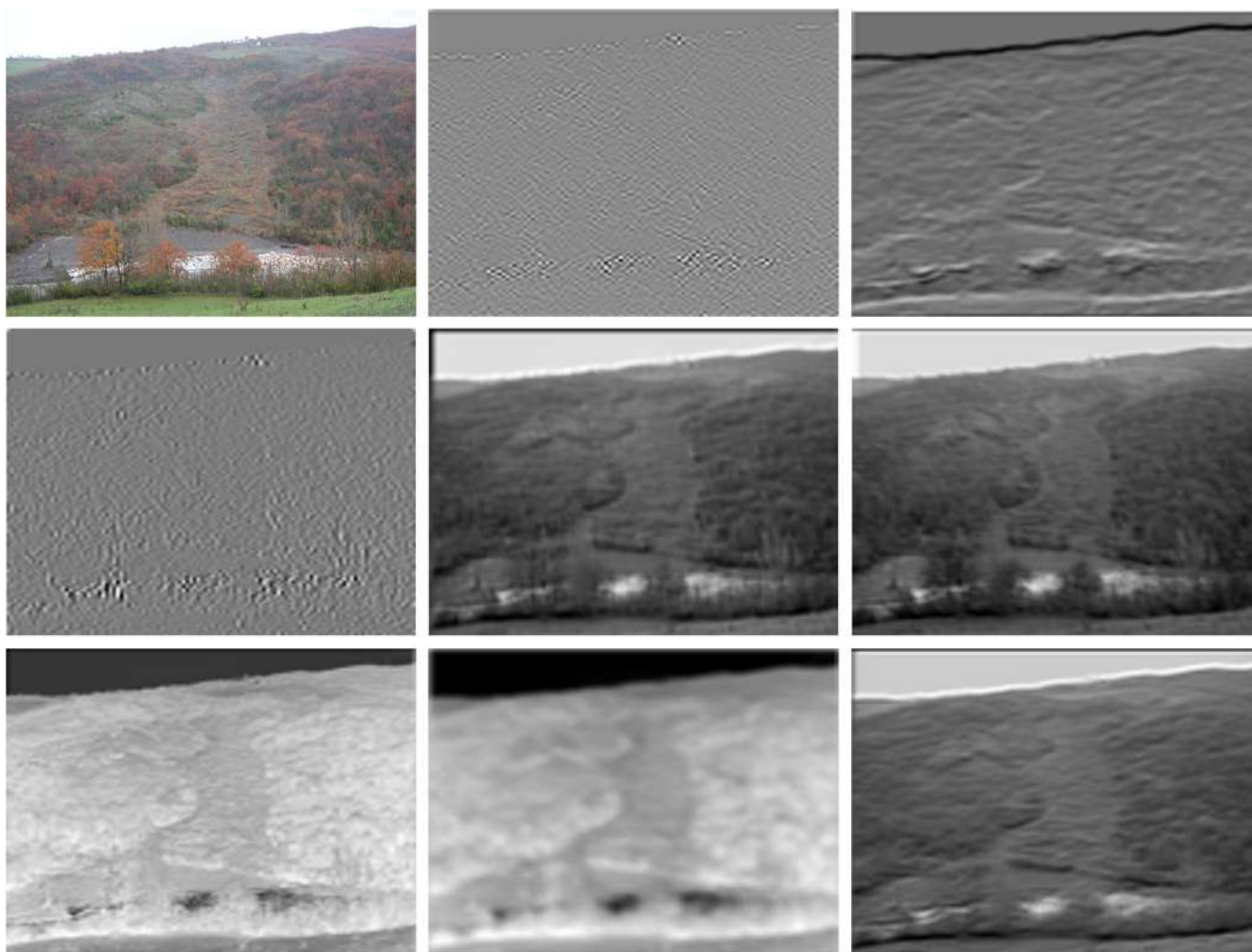
Fig. 13 Some activation of the image discussed in the text, generated by the first convolutional layer. The first image in the upper left corner is the original image passed to the network as input. It is clear that some activations are more relevant than others due to the fact that they are able to extract specific important features of the landform

Conclusions

A set of powerful convolutional neural networks publicly available have been adapted to recognize typical mass movement landforms within non-nadiral and non-standard pictures by transfer learning. The best parameter sets for the four tested algorithms have been determined by an iterative optimization procedure covering 576 different configurations. The accuracy and error analysis of such training runs shows that classification performances of such post-trained CNNs are consistently higher than those of the general-purpose original architectures and suitable for the usage in automated data mining of crowdsourced images. Furthermore, preliminary tests with basic and more advanced hardware configurations show that at least two of the optimal CNNs developed (Go-LanDLC and In-LanDLC) are compatible with usage in UAV and generic robot applications for automated survey and guidance, provided that some technical adjustments on image acquisition and pre-processing are made. A slight modification of the way the algorithm is applied may also allow for a quasi-real-time

Table 8 Degree of likelihood of class membership for the image in Fig. 13 for the four trained convolutional neural networks

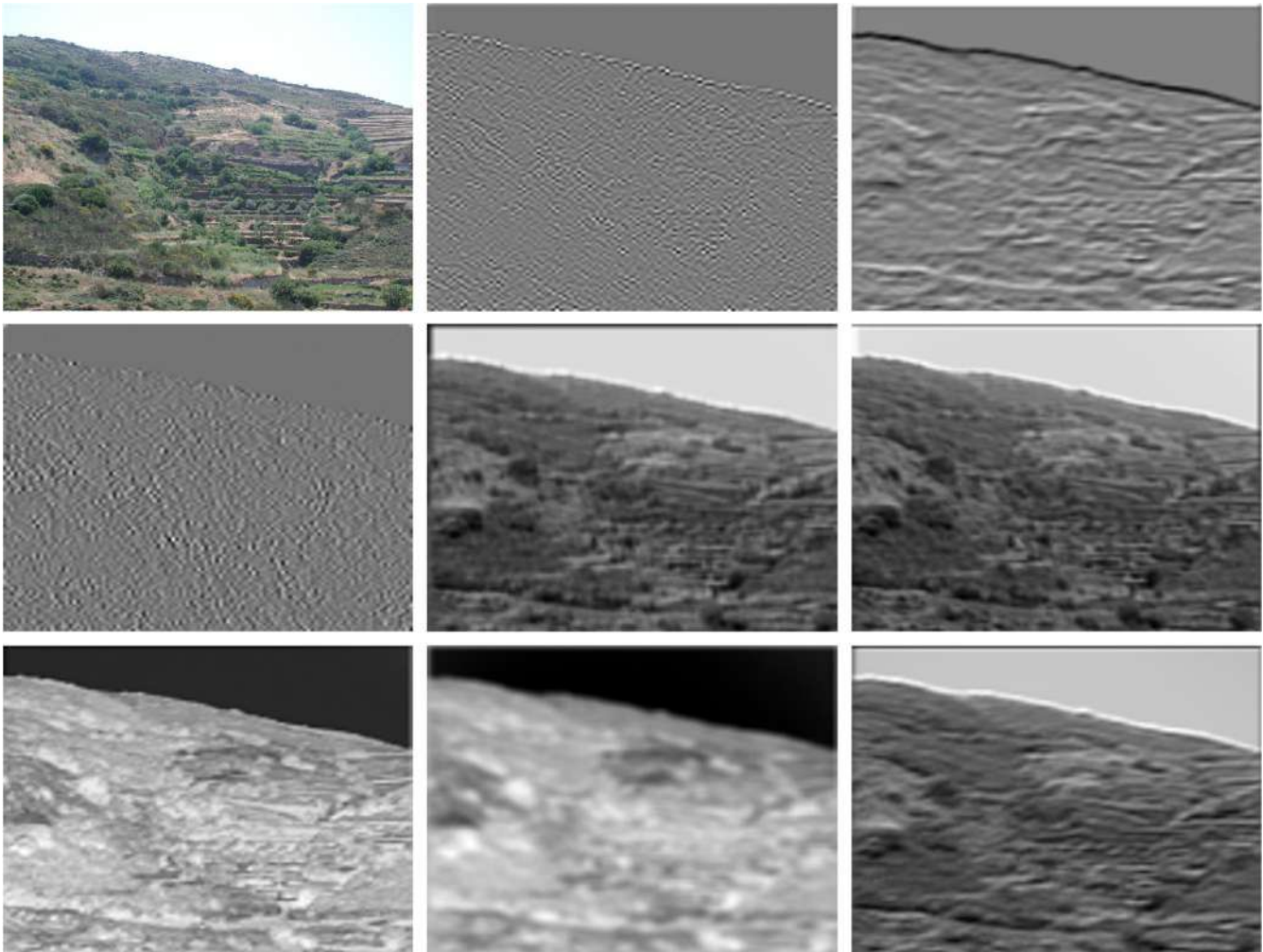|	Cliff	Alluvial fan	Landslide	Scree deposit	Stable slope
Go-LanDLC	0.000	0.003	0.997	0.000	0.000
GP-LanDLC	0.008	0.008	0.878	0.003	0.103
Re-LanDLC	0.060	0.022	0.149	0.021	0.748
In-LanDLC	0.003	0.002	0.711	0.001	0.283

Fig. 14 Example activations of a stable slope characterized by terracing and scattered vegetation that may render landform classification difficult. The activations have been generated by the first convolutional layer of the CNNs

scan of satellite VHR optical RGB images in a moving-window mode, thus potentially improving the capability of existing automated mapping tools. The four different versions of LanDLC are freely available for research purposes in the ONNX format under a CC BY NC 3.0 licence, as electronic supplementary material.

Further research is needed to work out the best trade-off between computational power on the one side and speed and compactness on the other, before developing actually implementable machine intelligence for automated landslide and landform detection. Among the priorities in future research

on this direction, we believe two will be essential: (i) a detailed analysis of a large number of convolutional schemes for the extraction of significant parameters for object recognition to reduce false negatives and false positives and (ii) the development of brand new CNNs specifically suited for landform recognition through full training with suitably large datasets (inexistent at present) of correctly labelled images. According to the present experience and to the work carried out for similar networks, we expect that such databases should have dimension in the order of 10^5–10^6 images.

Table 9 Degree of likelihood of class membership for the image in Fig. 14 for the four trained convolutional neural networks

	Cliff	Alluvial fan	Landslide	Scree deposit	Stable slope
Go-LanDLC	0.014	0.003	0.076	0.185	0.722
GP-LanDLC	0.028	0.006	0.158	0.026	0.782
Re-LanDLC	0.001	0.000	0.987	0.004	0.008
In-LanDLC	0.001	0.001	0.229	0.028	0.741

References

Allasia P, Baldo M, Giordan D, Godone D, Wrzesniak A, Lollino G (2019) Near real time monitoring systems and periodic surveys using a multi sensors UAV: the case of Ponzano landslide. In: Shakoor A, Cato K (eds) IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018 - Volume 1. Springer International Publishing, Cham, pp 303–310

Battistini A, Segoni S, Manzo G, Catani F, Casagli N (2013) Web data mining for automatic inventory of geohazards at national scale. Appl Geogr 43:147–158. https://doi.org/10.1016/j.apgeog.2013.06.012

Battistini A, Rosi A, Segoni S, Lagomarsino D, Catani F, Casagli N (2017) Validation of landslide hazard models using a semantic engine on online news. Appl Geogr 82:59–65. https://doi.org/10.1016/j.apgeog.2017.03.003

Batuwita R, Palade V (2012) Adjusted geometric-mean: a novel performance measure for imbalanced bioinformatics datasets learning. J Bioinforma Comput Biol 10:1250003. https://doi.org/10.1142/S0219720012500035

Bishop MP, James LA, Shroder JF, Walsh SJ (2012) Geospatial technologies and digital geomorphological mapping: concepts, issues and research. Geomorphology 137:5–26. https://doi.org/10.1016/j.geomorph.2011.06.027

Branco P, Torgo L, Ribeiro R (2015) A survey of predictive modelling under imbalanced distributions. ArXiv1505.01658 [cs.LG]

Catani F, Casagli N, Ermini L, Righini G, Menduni G (2005) Landslide hazard and risk mapping at catchment scale in the Arno River basin. Landslides 2:329–342

Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat Hazards Earth Syst Sci 13:2815–2831. https://doi.org/10.5194/nhess-13-2815-2013

Chae B-G, Park H-J, Catani F, Simoni A, Berti M (2017) Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. Geosci J 21:1033–1070. https://doi.org/10.1007/s12303-017-0034-4

Ciampalini A, Raspini F, Lagomarsino D, Catani F, Casagli N (2016) Landslide susceptibility map refinement using PSInSAR data. Remote Sens Environ 184:302–315. https://doi.org/10.1016/j.rse.2016.07.018

Corominas J, van Westen C, Frattini P, Cascini L, Malet J, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, Pitilakis K, Winter M, Pastor M, Ferlisi S, Tofani V, Hervás J, Smith J (2014) Recommendations for the quantitative analysis of landslide risk. Bull Eng Geol Environ 73(2):209–263. https://doi.org/10.1007/s10064-013-0538-8

Drăguţ L, Blaschke T (2006) Automated classification of landform elements using object-based image analysis. Geomorphology 81:330–344. https://doi.org/10.1016/j.geomorph.2006.04.013

Drăguţ L, Eisank C (2012) Automated object-based classification of topography from SRTM data. Geomorphology 141–142:21–33. https://doi.org/10.1016/j.geomorph.2011.12.001

Du L, You X, Li K, Meng L, Cheng G, Xiong L, Wang G (2019) Multi-modal deep learning for landform recognition. ISPRS J Photogramm Remote Sens 158:63–75. https://doi.org/10.1016/j.isprsjprs.2019.09.018

Eltner A, Kaiser A, Castillo C, Rock G, Neugirg F, Abellán A (2016) Image-based surface reconstruction in geomorphometry; merits, limits and developments. Earth Surf Dyn 4:359–389. https://doi.org/10.5194/esurf-4-359-2016

Ermini L, Catani F, Casagli N (2005) Artificial neural networks applied to landslide susceptibility assessment. Geomorphology 66(1–4):327–343. https://doi.org/10.1016/j.geomorph.2004.09.025

Evans IS (2012) Geomorphometry and landform mapping: what is a landform? Geomorphology 137:94–106. https://doi.org/10.1016/j.geomorph.2010.09.029

Ferri C, Hernández-Orallo J, Modroiu R (2009) An experimental comparison of performance measures for classification. Pattern Recogn Lett 30:27–38. https://doi.org/10.1016/j.patrec.2008.08.010

Forzieri G, Moser G, Catani F (2012) Assessment of hyperspectral MIVIS sensor capability for heterogeneous landscape classification. ISPRS J Photogramm Remote Sens 74:175–184. https://doi.org/10.1016/j.isprsjprs.2012.09.011

Forzieri G, Tanteri L, Moser G, Catani F (2013) Mapping natural and urban environments using airborne multi-sensor ADS40–MIVIS–LiDAR synergies. Int J Appl Earth Obs Geoinf 23(1):313–323. https://doi.org/10.1016/j.jag.2012.10.004

Giordan D, Manconi A, Tannant DD, Allasia P (2015) UAV: low-cost remote sensing for high-resolution investigation of landslides. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, Milan, pp 5344–5347

Giordan D, Hayakawa Y, Nex F, Remondino F, Tarolli P (2018) Review article: the use of remotely piloted aircraft systems (RPASs) for natural hazards monitoring and management. Nat Hazards Earth Syst Sci 18:1079–1096. https://doi.org/10.5194/nhess-18-1079-2018

He K, Zhang X, Ren S, Sun J (2015a) Delving deep into rectifiers: surpassing human-level performance on ImageNet Classification. Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, pp 1026–1034

He K, Zhang X, Ren S, Sun J (2015b) Deep residual learning for image recognition. ArXiv1512.03385 [cs.LG]

Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int J Uncertain Fuzziness Knowl-Based Syst 06:107–116. https://doi.org/10.1142/S0218488598000094

Hölbling D, Betts H, Spiekermann R, Phillips C (2016) Identifying spatio-temporal landslide hotspots on North Island, New Zealand, by analyzing historical and recent aerial photography. Geosciences 6:48. https://doi.org/10.3390/geosciences6040048

Huang Y, Yi S, Li Z, Shao S, Qin X (2011) Design of highway landslide warning and emergency response systems based on UAV. Proc. SPIE 8203, Remote Sensing of the Environment: The 17th China Conference on Remote Sensing, Hangzhou, p 820317. https://doi.org/10.1117/12.910424

Intrieri E, Bardi F, Fanti R, Gigli G, Fidolini F, Casagli N, Costanzo S, Raffo A, Di Massa G, Capparelli G, Versace P (2017) Big data managing in a landslide early warning system: experience from a ground-based interferometric radar application. Nat Hazards Earth Syst Sci 17:1713–1723. https://doi.org/10.5194/nhess-17-1713-2017

Jasiewicz J, Stepinski TF (2013) Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology 182:147–156. https://doi.org/10.1016/j.geomorph.2012.11.005

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds) Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp 1097–1105

Lee S, Ryu J-H, Won J-S, Park H-J (2004) Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Eng Geol 71:289–302. https://doi.org/10.1016/S0013-7952(03)00142-X

Lee J, Wang J, Crandall D, Sabanovic S, Fox G (2017) Real-time, cloud-based object detection for unmanned aerial vehicles. In: 2017 First IEEE International Conference on Robotic Computing (IRC). IEEE, Taichung, pp 36–43

Li H, Shi Y, Zhang B, Wang Y (2018) Superpixel-based feature for aerial image scene recognition. Sensors 18:156. https://doi.org/10.3390/s18010156

Liu Y, Wu L (2016) Geological disaster recognition on optical remote sensing images using deep learning. Procedia Comput Sci 91:566–575. https://doi.org/10.1016/j.procs.2016.07.144

Lu P, Stumpf A, Kerle N, Casagli N (2011) Object-oriented change detection for landslide rapid mapping. IEEE Geosci Remote Sens Lett 8:701–705. https://doi.org/10.1109/LGRS.2010.2101045

Lucieer A, de Jong SM, Turner D (2014) Mapping landslide displacements using Structure from Motion (SfM) and image correlation of multi-temporal UAV photography. Prog Phys Geogr Earth Environ 38:97–116. https://doi.org/10.1177/0309133313515293

Minaeian S, Liu J, Son Y-J (2016) Vision-based target detection and localization via a team of cooperative UAV and UGVs. IEEE Trans Syst Man Cybern Syst 46:1005–1016. https://doi.org/10.1109/TSMC.2015.2491878

Mondini AC, Guzzetti F, Reichenbach P, Rossi M, Cardinali M, Ardizzone F (2011) Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. Remote Sens Environ 115:1743–1757. https://doi.org/10.1016/j.rse.2011.03.006

Niesterowicz J, Stepinski TF (2013) Regionalization of multi-categorical landscapes using machine vision methods. Appl Geogr 45:250–258. https://doi.org/10.1016/j.apgeog.2013.09.023

Niethammer U, James MR, Rothmund S, Travelletti J, Joswig M (2012) UAV-based remote sensing of the Super-Sauze landslide: evaluation and results. Eng Geol 128:2–11. https://doi.org/10.1016/j.enggeo.2011.03.012

Pradhan B, Lee S (2010) Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. Landslides 7:13–30. https://doi.org/10.1007/s10346-009-0183-2

Rossi G, Tanteri L, Tofani V, Vannocci P, Moretti S, Casagli N (2018) Multitemporal UAV surveys for landslide mapping and characterization. Landslides 15:1045–1052. https://doi.org/10.1007/s10346-018-0978-0

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. Int J Comput Vis 115:211–252. https://doi.org/10.1007/s11263-015-0816-y

Scaioni M, Longoni L, Melillo V, Papini M (2014) Remote sensing for landslide investigations: an overview of recent achievements and perspectives. Remote Sens 6:9600–9652. https://doi.org/10.3390/rs6109600

Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35:1285–1298. https://doi.org/10.1109/TMI.2016.2528162

Smith L, Liang Q, James P, Lin W (2017) Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework: assessing the utility of social media for flood risk management. J Flood Risk Manag 10:370–380. https://doi.org/10.1111/jfr3.12154

Stumpf A, Kerle N (2011) Object-oriented mapping of landslides using Random Forests. Remote Sens Environ 115:2564–2577. https://doi.org/10.1016/j.rse.2011.05.013

Sun Y, Wong AKC, Kamel MS (2009) Classification of imbalanced data: a review. Int J Pattern Recognit Artif Intell 23:687–719. https://doi.org/10.1142/S0218001409007326

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015a) Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, pp 1–9

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2015b) Rethinking the inception architecture for computer vision. ArXiv1512.00567 [cs.CV]

Tarolli P (2014) High-resolution topography for understanding Earth surface processes: opportunities and challenges. Geomorphology 216:295–312. https://doi.org/10.1016/j.geomorph.2014.03.008

Turner D, Lucieer A, de Jong S (2015) Time series analysis of landslide dynamics using an unmanned aerial vehicle (UAV). Remote Sens 7:1736–1757. https://doi.org/10.3390/rs70201736

Uzielli M, Catani F, Tofani V, Casagli N (2015a) Risk analysis for the Ancona landslide—I: characterization of landslide kinematics. Landslides 12:69–82. https://doi.org/10.1007/s10346-014-0474-0

Uzielli M, Catani F, Tofani V, Casagli N (2015b) Risk analysis for the Ancona landslide—II: estimation of risk to buildings. Landslides 12(1):83–100. https://doi.org/10.1007/s10346-014-0477-x

Yilmaz I (2010) The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks. Environ Earth Sci 60:505–519. https://doi.org/10.1007/s12665-009-0191-5

Zhou B, Khosla A, Lapedriza A, Torralba A, Oliva A (2016) Places: an image database for deep scene understanding. ArXiv1610.02055 [cs.CV]

Zhou C, Yin K, Cao Y, Ahmed B, Li Y, Catani F, Pourghasemi HR (2018a) Landslide susceptibility modeling applying machine learning methods: a case study from Longju in the Three Gorges Reservoir area, China. Comput Geosci 112:23–37. https://doi.org/10.1016/j.cageo.2017.11.019

Zhou C, Yin K, Cao Y, Intrieri E, Ahmed B, Catani F (2018b) Displacement prediction of step-like landslide by applying a novel kernel extreme learning machine method. Landslides 15:2211–2225. https://doi.org/10.1007/s10346-018-1022-0

F. Catani (✉)
Earth Sciences Department,
University of Florence,
Via La Pira, 4, 50121, Florence, Italy
Email: filippo.catani@unifi.it